Port Freight Statistics – User Engagement Exercise

1 Contents

2 Introduction

2.1.1 **The Department is reviewing the methods used to produce *Port Freight Statistics*[1]** and will be working with users and stakeholders to ensure that they are useful and accurate and continue to be fit for purpose. *Port Freight Statistics* are an annual publication produced by the Department for Transport on trends and patterns in the handling of freight traffic at UK sea ports.

2.1.2 By continuously improving the series, the Department can ensure that it meets the requirements of government, public services, business, researchers and the public.

2.1.3 *The Code of Practice for Official Statistics*[2] provides a framework for the review. It sets out the protocols and practices that should be followed for the production and publication of statistics, including meeting user needs, use of sound methods and robust quality assurance.

2.1.4 Through this review, the Department aims to:

- Consult informally with users and stakeholders to determine their requirements;
- Review the publication's existing methodology;
- Determine potential alternative data sources and methodologies, and investigate their feasibility and accuracy;
- Assess and compare each option; and,
- Determine the most suitable options for future publications.

[1] https://www.gov.uk/government/collections/maritime-and-shipping-statistics
[2] https://www.statisticsauthority.gov.uk/monitoring-and-assessment/code-of-practice/

3 Background

3.1.1 The maritime sector is essential to the UK economy, contributing at least £13 billion each year[3]. With around 95% of UK imported and exported goods transported by sea (including approximately 40% of our food[4] and about one quarter of our energy[5]) it is clear that our island nation has a huge reliance on the shipping industry.

3.1.2 On top of this, an estimated 23,000 UK nationals were seafarers working regularly at sea in 2015[6], whilst UK ports handled almost 500 million tonnes of freight in 2015[7] - this includes large quantities of goods from high value sectors, for example, 130 million tonnes of bulk fuels and 9 million tonnes of agricultural goods were brought into the UK in 2015. It is therefore important that there is an accurate data collection system in place to monitor the volume of goods shipped into and out of the UK as there is a wide range of users reliant on such data.

3.1.3 The Department for Transport (DfT) has a requirement to report UK *Port Freight Statistics* to the European Commission, under Directive 2009/42/EC, which requires comparable, reliable, synchronised and regular statistical data on the scale and development of the carriage of goods and passengers by sea.

3.1.4 It is also important that DfT collects these figures to meet the needs of a wide range of other users, including researchers, academics, and other government departments. These statistics provide the vital evidence needed for decision making and enable users to have a detailed picture of how the maritime sector is behaving and changing. For example, without accurate statistics on the volume of goods entering or leaving UK ports, it would be difficult to say whether the UK maritime sector is growing or not, and it could be difficult to produce contingency plans for the event of disruption at a key port.

3.1.5 There are many other important uses of *Port Freight Statistics*, including the development and validation of port traffic forecasts[8] and use in market analysis by transport consultants and businesses.

3.1.6 In order to satisfy each of these uses and requirements, DfT produces four quarterly and one annual publication on UK Port Freight. Each of these publications is badged as *National Statistics*, and as such DfT, as the producers of these statistics, seeks to achieve continuous improvement and ensure that they meet the requirements of government, public services, business, researchers and the public. The Code of Practice for Official Statistics sets out the protocols and practices that are to be followed, including meeting user needs, a high standard of production and management, and that statistics are well explained. This review will address these practices and protocols and assesses the current methodology and possible alternatives.

3.1.7 It is acknowledged by the Department that there are known issues with the current data collection methods, which place burden upon shipping agents and operators. It is also challenging to maintain complete coverage of traffic at UK ports and a high level of quality for all data collected due to the extensive volume of data collected.

3.1.8 This Department's review of *Port Freight Statistics* is an opportunity for the Department to assess the feasibility of new methods of data collection, and to investigate the potential use of different data sources, in order to improve the way data are collected, enabling a more efficient process.

[3] DfT analysis of ONS GVA from the Input-Output Supply and Use Tables, apportioned using detailed SIC data for approximate GVA from the ONS Annual Business Survey

[4] UK Food Production to Supply Ratio ('Self-Sufficiency') all food, Agriculture in the United Kingdom, 2014, Department for Environment, Food and Rural Affairs https://www.gov.uk/government/uploads/system/uploads/attachment_data/ file/430411/auk-2014-28may15a.pdf

[5] Digest of UK Energy Statistics, 2014, Department of Energy and Climate Change estimates that in 2013, net imports accounted for 47 per cent of energy used in the UK. https://www.gov.uk/government/uploads/system/uploads/ attachment_data/file/338750/DUKES_2014_printed.pdf Applying some basic assumptions (that all coal imports are imported by sea, that half of all petroleum imports are imported by sea and 55 per cent of LNG imports come into the country by pipeline) therefore implies that approximately 28 per cent of UK energy used in the UK arrives in the UK by sea

[6] Seafarers Statistics 2015, DfT

[7] Port Freight Statistics 2015, DfT

[8] UK Port Demand Forecasts to 2030, 2006 MDS Transmodal Ltd. http://webarchive.nationalarchives.gov.uk/+/http:/www.dft.gov.uk/consultations/archive/2006/ppr/ukportdemandforecaststo2030.pdf

4 Feedback

4.1.1 In order to establish both user and stakeholder needs and to understand the burden placed on providers when supplying data for *Port Freight Statistics*, the Department is interested in hearing your views. Section 9 invites users to respond to a number of specific questions relating to the data collection process. Please respond to these questions by completing the online survey or contacting the Department directly using the contact details provided.

5 Content and structure of the publications

5.1.1 The main statistical release includes a combination of charts and commentary for each of the following topics:

- Key overall trends,
- Traffic by direction,
- Comparisons with other sources,
- Traffic by Cargo category
- Traffic by Port
- Trade routes
- Trade with EU countries
- Number of Arrivals

5.1.2 At the end of main release document, there are sections on user feedback, strengths and weaknesses of the data, links to further information, and a glossary.

5.1.3 The separate technical notes provide more detail on the data sources, methodology, assumptions and definitions.

5.1.4 Users can also access summary tables of data covering the topics listed above. Previously, these were published in Excel format, but in future releases they will be provided in OpenDocument Spreadsheet (ODS) format, in line with guidance[9] issued by Cabinet Office.

5.1.5 Documents can be accessed on the Department's website at the following web addresses:

- Statistical release documents and factsheets:
https://www.gov.uk/government/collections/maritime-and-shipping-statistics

- Summary data tables:
  - UK ports and traffic (PORT01)
https://www.gov.uk/government/statistical-data-sets/port01-uk-ports-and-traffic
  - Port freight (PORT02)
https://www.gov.uk/government/statistical-data-sets/port02-freight
  - Key port statistics (PORT03)
https://www.gov.uk/government/statistical-data-sets/port03-key-port-statistics
  - Individual port traffic (PORT04)
https://www.gov.uk/government/statistical-data-sets/port04-individual-port-traffic
  - Traffic, indices and quarterly breakdowns (PORT05)

[9] More information about this guidance can be found here:
https://www.gov.uk/government/news/open-document-formats-selected-to-meet-user-needs

https://www.gov.uk/government/statistical-data-sets/port05-traffic-indices-and-quarterly-breakdowns

- UK ship arrivals (PORT06)
https://www.gov.uk/government/statistical-data-sets/port06-uk-ship-arrivals

- Technical notes and documents:
https://www.gov.uk/government/publications/maritime-and-shipping-statistics-guidance

- User guides for data suppliers:
https://www.gov.uk/government/publications/maritime-statistics-directive-reporting-guidance

5.1.6 The Department is particularly interested to understand which documents users are making frequent use of and is interested in any comments on these documents.

Questions to users: Publication content and structure

- Are there specific topics in the publication which you find useful?
- Do you frequently use any data tables or any of the other documents available?
- Are there any additional outputs that the Department should consider producing?

6 Current Methodology

6.1.1 This section sets out the current methods used for *Port Freight Statistics,* and the processing adjustments applied to this data to produce the annual results. Annex B – Data collection overview, gives an overview of the data collection process as a whole, including each stage of data validation and grossing and shows how they fit in the overall data collection process map.

6.2 Data sources

6.2.1 Detailed freight information is collected from shipping lines, operators and shipping agents, at least every quarter – in practice some prefer to provide information on a daily basis. The detailed route and ship flag information required is only generally available from lines, operators and agents. Ports supply more limited information on a quarterly and annual basis, which is used to align totals and to publish more timely provisional statistics.

6.3 Data collection forms and their requirements

6.3.1 Six different data collection forms are used to collect the required information from shipping lines, shipping agents and ports. The following list gives an overview of what is collected on each form:

- MSD1 – Freight movements return by shipping lines/agents (at least quarterly)
- MSD2 – Quarterly port traffic return of freight tonnages by major ports
- MSD2X – Annual port traffic return of detailed freight tonnages by major ports
- MSD3 – Quarterly port return of shipping lines/agents by major ports
- MSD4 – Quarterly vessel return by major ports
- MSD5 – Annual port traffic return of freight tonnages by minor ports

6.3.2 These forms are returned using four different methods:

- An online system called the Internet Statistics Data Entry System (iSDES)
- Through the supply of an ASCII flat file
- Using a Generic Statistical Message (GESMES) – this is a standard method used in Electronic Data Exchange
- Or, in rare cases, using paper forms

6.3.3 All three electronic methods of data receipt provide some level of built in validation, such as checks for required data fields, values outside of a specified range and non-existent category codes. Depending on the system used, these checks are either flagged as they are entered into the system (iSDES) or result in an automatic rejection of the data. Each year

approximately 1 in 10 submissions contain an error and the Department would like to understand how this number might be reduced, in order to improve efficiency. This could be through a number of options as detailed in Section 6.4.

Questions to users: Data collection forms

- Are data suppliers happy with the current options for providing data to DfT?
- If the Department hosted workshops or training with the aim of working with suppliers to reduce the number of submissions with errors, would this be of interest?

6.4 Processing adjustments

6.4.1 The statistics series is badged as a National Statistic, meaning that it has been assessed by the UK Statistics Authority and was found to be compliant with the Code of Practice for Official Statistics. It is therefore deemed to be trustworthy, high quality and of public value. However, the data sources have some limitations and the review will re-evaluate these and aim to address them where possible.

6.4.2 **Data validation and assurance** – Current data collection methods rely heavily on human intervention in order to identify and correct errors. For example, if a port supplies a data return for a ship arriving at that port, but there is no corresponding data return for that ship from the shipping agent, this is currently identified and investigated manually.

6.4.3 The relevant port(s) is contacted to establish who the shipping agent was, assuming that the port is able to provide this information – in some cases they are not. The agent is then contacted to supply a data return, which would be manually validated against the detailed cargo return supplied by the port, to provide quality assurance. This circumstance may occur many times for a single port, requiring considerable resource to investigate and making the quality assurance process difficult to replicate each year, as human inference is often a key component in the discovery of errors.

6.4.4 The data collection exercise places considerable burden on the businesses required to supply data. Almost two thirds of all returns are supplied by a Port Community System - a company who has been paid to manage and supply all data returns – which shipping agents may see as a way of reducing this burden.

6.4.5 It is a common outcome that even after this manual validation, the port and agent data does not entirely match, indicating that there are still differences between the sources. It would be resource intensive to investigate every difference indefinitely until all data matches, and in most cases this would not be possible, thus another solution is required.

6.4.6 **Data grossing** - In order to reach a solution, the annual port traffic data are used to adjust the information supplied by shipping lines and agents, i.e. to

make an adjustment to the data to correct for any missing or misclassified data from shipping lines/agents. Annual port traffic data from the ports is deemed more likely to be complete and accurate simply because there are far fewer major ports to report data than there are shipping lines and agents.

6.4.7 For an example of the data grossing process please see Annex A – Example of validation and grossing methods.

6.4.8 This process of data grossing is again highly dependent on both the success of follow-up work to improve the overall quality of data supplied to the Department, and on the initial quality of the data supplied by ports and agents. This is a challenge because both processes are largely manual, relying on human intervention, which makes it difficult to maintain consistent levels of data quality.

6.4.9 This is compounded because the vast majority of this validation and grossing takes place between 6 and 18 months after the vessel movements took place. This means that shipping lines/agents or ports often have to consult their own records to investigate any queries, which relies on them collecting and maintaining good records.

6.4.10 All queries raised by the Department are raised within a short period of time, thus increasing the burden on data suppliers. If a “live” validation system (i.e. a system which is able to look at other data returns or other data sources, and flag missing data) was in place to flag any issues with data suppliers throughout the year, this could reduce burden and also make the validation process more efficient.

6.4.11 The Department is using the review as an opportunity to investigate other data sources – such as Automatic Identification System (AIS) data and the UK Maritime National Single Window (MNSW) – which may offer solutions for improving the processing methods of *Port Freight Statistics* in the future.

Questions to users: Processing adjustments

- Would “live” validation be suitable?
- How might the resultant more timely and frequent checks on data submissions affect the level of burden for data suppliers?
- Would the earlier identification and correction of errors in data submissions be valuable to data suppliers?

7 Alternative data sources

7.1.1 One potential way of improving the current *Port Freight Statistics* data collection system is to look at alternative data sources. It is possible that the data required are already collected by another government agency or department. If DfT were to obtain this data from them, the burden upon data suppliers could be reduced, and extra resource could be utilised for further data validation.

7.1.2 There are four different data sources identified here as having potential to improve the current collection methods:

1. UK Maritime National Single Window (MNSW)
2. Automatic Identification System (AIS) data
3. Her Majesty's Revenue and Customs (HMRC) Trade data
4. The Consolidated European Reporting System (CERS)

7.1.3 For a summary of these sources and how they meet *Port Freight Statistics* requirements, please see Section 7.6.

7.2 UK Maritime National Single Window

7.2.1 It is a requirement of Directive 2010/65/EU that EU Member States seek to reduce the administrative burdens for shipping companies, harmonise and simplify data collection procedures, and gather information more efficiently. With this in mind, each EU Member State is currently developing a MNSW system to facilitate electronic data collection of the information required across several Directives and Regulations.

7.2.2 The launch of the UK MNSW introduces an online portal through which ships' pre-arrival reports can be submitted to certain authorities including the UK government. During the pilot phase, the MNSW will only accept:

- FAL form 1 ('IMO general declaration')
- FAL form 5 ('Crew list')
- FAL form 6 ('Passenger list')

7.2.3 The forms provide advance data regarding the ship, its voyage, stores, crew, passengers, dangerous cargo, and security, waste/health information. The UK MNSW is currently in its pilot phase, meaning that only Border Force and Her Majesty's Revenue and Customs have access to this data at present.

7.3 Automatic Identification System data

7.3.1 Regulation 19 of Safety of Life at Sea (SOLAS) Chapter V - Carriage requirements for shipborne navigational systems and equipment, sets out navigational equipment to be carried on board ships, according to ship type. In 2000 the International Maritime Organisation (IMO) adopted a new requirement (as part of a revised chapter V) for all ships to carry an AIS capable of providing information about the ship, to other ships and coastal authorities, automatically.

7.3.2 The regulation requires an AIS to be fitted aboard all:

- cargo ships of 300 gross tonnes and upwards engaged on international voyages
- cargo ships of 500 gross tonnes and upwards not engaged on international voyages
- passenger ships irrespective of size.

7.3.3 The regulation requires that AIS shall:

- provide information - including the ship's identity, type, position, course, speed, navigational status and other safety-related information - automatically to appropriately equipped shore stations, other ships and aircraft;
- receive automatically such information from similarly fitted ships;
- monitor and track ships;
- exchange data with shore-based facilities.

7.3.4 AIS data are supplied to the Maritime and Coastguard Agency (MCA) and DfT are currently investigating accessing this data. There is potential to use AIS data to improve data quality by using AIS data to validate data returns – i.e. to check load types and weights, to check voyage dates, to check ship details are correct and to check for any missing data

7.3.5 Information regarding the cargo carried by a ship is not transmitted via AIS, in particular the cargo type and weight are not transmitted. It is therefore unlikely that AIS data could be used to meet all of the requirements of the regulations on *Port Freight statistics*. Data will be investigated further if obtained from the MCA.

7.4 HMRC Trade data

7.4.1 Trade Statistics data are available on the uktradeinfo website which is managed by HMRC. These statistics record the movement of goods between the UK and both EU and non-EU countries, for trade purposes, and are collected through a system called Intrastat. Detailed documentation of this system is available on the uktradeinfo website.

7.4.2 Data for EU trade are collected through a monthly survey of all businesses with annual arrivals over £1.5m or dispatches over £250,000. For businesses under this threshold data are estimated. Data for non-EU trade are collected through customs declarations.

7.4.3 The fields that are collected are as follows:

- Commodity Code
- Invoice Value
- Net mass (kg)/Supplementary Unit (where appropriate as determined by Commodity Code)
- Country of Destination or Dispatch (COD)
- Delivery Terms (if the business reaches the Delivery terms threshold)
- Nature of Transaction

7.4.4 As detailed in Table 1, the coverage of this data source is not complete enough to be considered a reasonable replacement for the current collection system. The main information which is not covered by this source is information regarding the ship carrying the goods – this is essential to verify data returns and to add supplementary information to the dataset. Domestic traffic is also not collected in these statistics. There is some potential for trade statistics data to be used to validate other data, but without complete coverage of domestic traffic, and ship information, this source is not entirely suitable.

7.4.5 There are also other areas of coverage which, although less critical, are still lacking from this data source, for example it does not cover goods for loan or temporary use, trade by private persons, or goods supplied free of charge. These are all covered in the current data collection system.

7.5 Consolidated European Reporting System

7.5.1 The MCA collects data on arrivals and departures in its CERS database. Arrival port notifications must be entered onto CERS for:

- Every sea-going merchant vessel (equal to or greater than 300 GT) arriving at a UK port.
- Any sea-going merchant vessel (irrespective of size) arriving at a UK port from outside the EU and carrying dangerous or polluting goods.

7.5.2 The operator, agent or master of such a ship must provide the required port arrival notification to the port authority at least 24 hours in advance of ship arrival, or at the latest at the time the ship leaves the previous port, if the voyage time is less than 24 hours; or if the port of arrival is not known or it is changed during the voyage, as soon as this information is available.

7.5.3 The port authority shall, upon receipt, input the port arrival notification received from the operator, agent or master onto CERS.

7.5.4 CERS does not require any information regarding the goods carried by a vessel to be input, unless they are classified as dangerous goods. It is therefore unlikely that data from CERS would meet the requirements of *Port Freight Statistics*. However, as with AIS data, CERS data may be used to validate these statistics. The Department is currently investigating accessing

CERS data and this will be investigated further upon receipt of data from the MCA.

7.5.5 There is no information regarding the coverage, quality or suitability of CERS data available at present. The suitability of CERS data would be investigated as and when data are obtained.

7.6 Summary of alternative data sources

7.6.1 The following table sets out the coverage of each of these sources in terms of the key data requirements for *Port Freight Statistics*:

Table 1: Suitability of data sources for meeting the key requirements of Port Freight Statistics

	Current method	Data Source			
	***Port Freight Statistics* data collection**	**UKMNSW**	**AIS**	**HMRC**	**CERS**
Timeliness	Weekly, monthly and annual data	Unknown, but transmissions are frequent	Unknown, but transmissions are frequent	14 days - 1 month	Unknown, but transmissions are frequent
Coverage:					
MO number	✓	✓	✓	~	✓
Goods type info.	✓	✓		✓	
Tonnage info.	✓	✓		✓	
Unit info.	✓				
Date of voyage/arrival	✓	✓	✓	✓	✓
Agent/Operator info.	✓	✓	✓	~	~
Last port of call	✓	✓	✓	Country only	✓
Arrival port	✓	✓	✓	✓	✓
All required goods types	✓	~		~	
Minor port coverage	✓	✓	✓	~	~
Domestic traffic	✓	~	✓		~
International traffic	✓	✓	✓	✓	✓

~ symbol indicates unknown or incomplete coverage

7.6.2 Table 1 shows that broadly, the UK MNSW is the closest to meeting all of the required data fields for *Port Freight Statistics*, however this data source is currently only in its pilot phase and data are not yet available to the Department. This source of data will be investigated further if and when it becomes available as this could be considered a feasible replacement for some parts of the current data collection system.

7.6.3 There is potential for AIS data to be used to improve the data validation processes, as previously noted, and DfT are investigating this further. It is worth noting that due to a lack of information regarding goods carried, this system is unlikely to fully meet the requirements of *Port Freight Statistics*.

7.6.4 Trade data could be used either to improve data validation techniques, or as a possible replacement for international data. However, at present there are too many unknowns with regard to the completeness of the data and further investigation of this data source is required to understand whether it could fully meet our requirements. A high-level comparison of *Port Freight Statistics* and international trade data from HMRC is available in the 2015 *Port Freight Statistics* publication, with further discussion in the Notes and Definitions document.

7.6.5 Of the other data sources covered here, CERS data appears to have similar potential to AIS data with regard to improving data validation methods. However there are further unknowns with regard to coverage, so it is unlikely that this data source will be investigated further as a priority.

7.6.6 In Section 9 your views are sought on whether there are other sources of data that you are aware of which may meet these requirements, and how the potential use of these data sources might reduce burden for data suppliers or improve the outputs of *Port Freight Statistics*.

Questions to users: Alternative data sources

- Do you agree with the Department's assessment of the alternative sources listed in this document?
- Are there any other data sources you would like the Department to consider using for *Port Freight Statistics*.

7.7 Limitations

7.7.1 A potential drawback to using other data sources such as these is that they would be outside of the Department's direct control, meaning that the Department would not have a direct say in the data collection process and the methods used. This could lead to breaks in the *Port Freight Statistics* time series, or even the complete loss of data. This is a risk which would need to be assessed and mitigated before making any changes to the current data collection process.

7.8 Further exploration

7.8.1 The Department acknowledges that the requirement to supply Port Freight data may create a large amount of burden on operators and agents, however as outlined earlier this information is important to the Department. It is also acknowledged that providing this information (or parts of this information) to

multiple organisations creates unnecessary burden, which if possible DfT aims to reduce.

Questions to users: Alternative data sources

- Do you supply the same or similar information to other organisations? If so, who?
- If the Department collected this information directly from another organisation, would this be beneficial to you as a data supplier?
- Could you provide other documentation such as Bills of Lading directly to the Department?

7.9 International data collection

7.9.1 Looking at the methods used by other European countries who are also required to meet the criteria set out by Directive 2009/42/EC, most countries collect data in a similar way to DfT.

7.9.2 A total of 21 other European countries collect data to compile their own *Port Freight Statistics* and supply the necessary outputs to the European Commission. The vast majority of these countries collect data through ports and/or shipping agents, however there a few countries which collect their data using notably different methods.

7.9.3 Estonia and Croatia for example have a system in place similar to the UK MNSW, which as discussed is currently in pilot stage and will be explored further in the near future.

7.9.4 The other countries which collect data in a different way are Germany and The Netherlands, who combine data from ports with that from customs authorities in order to meet the requirements of the Directive. As discussed, this option could be explored by DfT, and may be supported by experience and knowledge of German and Dutch counterparts.

7.9.5 Comparisons with data from countries who collect data from customs authorities highlighted potential issues which could arise if data collection methods were changed. The details from these comparisons are international available in the 2015 *Port Freight Statistics* publication, with further discussion in the Notes and Definitions document.

Questions to users: Comparisons with other sources

- Is the comparison with other data sources useful to users?

7.9.6 Outgoing data from The Netherlands, for Port Freight traffic to the UK, is considerably lower than UK data for traffic from the Netherlands. This is not the case for data in the opposite direction, which may suggest that customs authorities in some countries are under recording outgoing data. Further investigation would be required before moving to using data gathered by any other organisation or Department.

8 Alternative sampling approaches

8.1.1 Currently, *Port Freight Statistics* are produced using what is effectively a census of all port freight traffic to or from the UK. Another approach explored within the review is the use of different survey sampling techniques. In this section this approach is explored, as are details of its strengths and limitations.

8.1.2 The main aim of using sampling would be to reduce the overall burden upon data suppliers, i.e. less resource would be required to supply and validate data for 25% of traffic, compared with all traffic, whilst maintaining an appropriate level of accuracy of *Port Freight Statistics*.

8.2 Identifying the population

8.2.1 In order to select a sample, knowledge of the entire population of vessels is required. This is held in the form of data from the IHS Global international vessel database and the Maritime and Coastguard Agency's UK Ship Register. Note that vessels are selected as the population, as opposed to vessel movements, as there would be no knowledge of vessel movements until data for all movements are collected, which would then making sampling redundant. For example, the Department receives data on vessel arrivals from Lloyds List Intelligence 3 months after the end of the calendar year, at which point it would be too late to use to sample January to March of the same year.

8.2.2 It would also be possible to select a sample by defining the population as all UK shipping agents or all UK major ports, however neither of these approaches would be suitable as it could become confusing for ports/agents who are not asked to provide data in one year, if they are then asked to do so in a following year, but they do not have the resource or knowledge of how to do so.

8.2.3 Having no information for a specific port would require the estimation of data for that port, which fail to identify changes in trends. This information is key to how *Port Freight Statistics* are used in the Department and hence it is a better approach to sample by vessel, using the vessel's IMO number.

8.2.4 This method is not without its issues in that any sample must be selected prior to data collection - before January of the current data collection year. In order to do this it would only be possible to use a database relating to the previous year and so any new vessels in operation would have zero probability of selection. Also, any vessels scrapped at the end of the previous year would have a probability of selection when they would not be part of the population.

8.3 Simple random sampling

8.3.1 To begin by taking the simplest approach, we can look at whether simple random sampling is suitable. This method involves selecting an entirely random sample from the defined population – in this case all vessels carrying

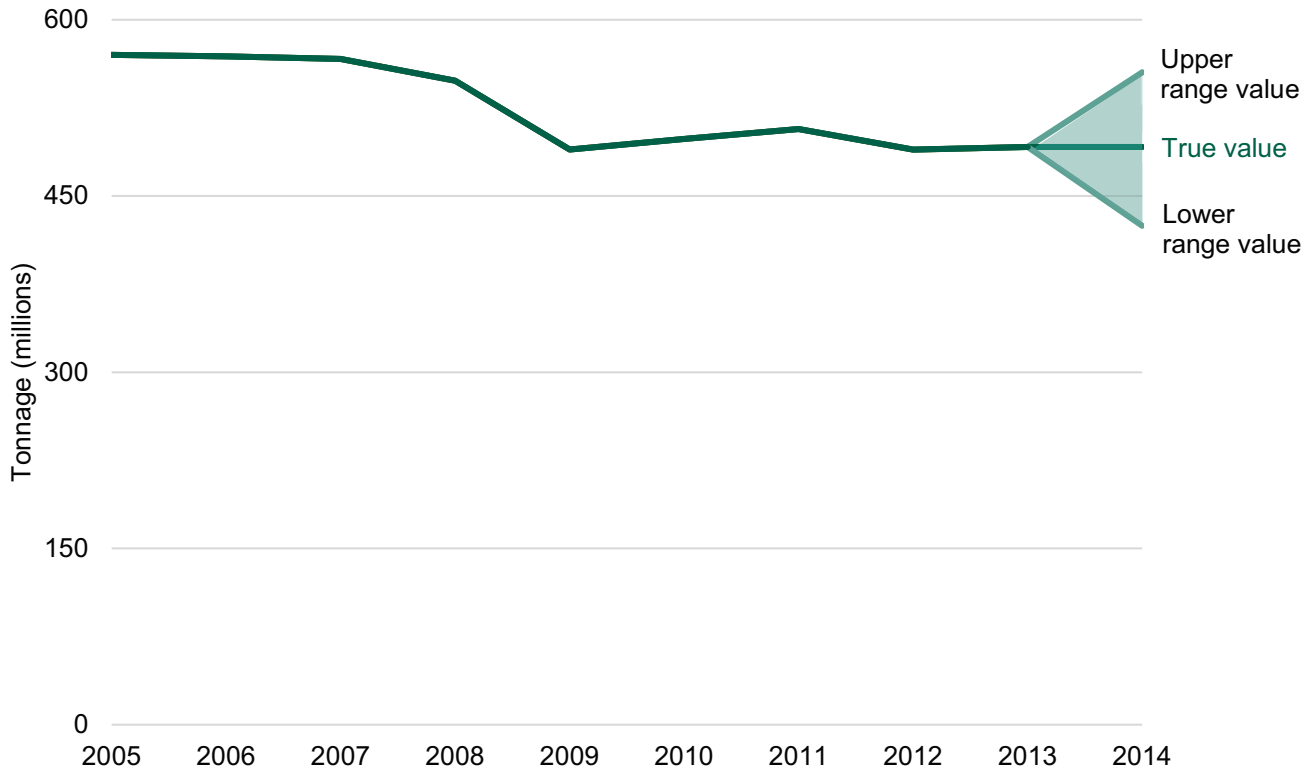
port freight traffic at major UK ports – where each member (or vessel) of the population has an equal and non-zero probability of selection.

8.3.2 To test this approach multiple samples of vessels are selected, for different proportions of the population. The data are then retrospectively examined for those vessels to assess the accuracy of the sample in estimating the total tonnage for that year.

8.3.3 Several different proportions of the population were tested, with the general principle that a smaller sample would be easier and require less resource to collect, but would also be less accurate in terms of estimating the population mean – that is the total tonnage for UK port freight.

8.3.4 For 2014 data, 25 samples of 25% of all vessels were selected and sample means were calculated. These means had extremely large variation – they ranged from 425 million to 555 million tonnes, a range of 130 million tonnes. This represents +/- 13% of the population mean, which would be considered a very large year on year change in tonnage (larger than any single year change recorded). It is for this reason that simple random sampling is not considered a viable method.

Figure 1: The range of sampling UK major port tonnage, 2005-2014



8.3.5 It would be possible to further increase the sample proportion in order to increase the accuracy of the sample mean, but at this point the savings from increased efficiency and reduced burden would be too greatly reduced. There are other approaches which can be explored to better result.

8.3.6 The main challenge with sampling is a result of the extremely small proportion of vessels which handle UK freight. Less than 5% of all vessels in a year handle UK freight, meaning each vessel has a relatively low probability of selection when sampling. This creates large variation in the overall sample mean. Other sampling methods were explored, but no suitable method was found which could overcome this problem.

8.3.7 Therefore the Department concludes that sampling would not be an appropriate method to explore any further, but welcomes any comments that users have on this approach, as well as suggestions for how it might be adopted successfully.

Questions to users: Alternative sampling approaches

- Do you agree with the Department's assessment of sampling as an approach to data collection?
- Are there any further approaches to sampling which you feel that the Department should explore?

9 How to respond

9.1.1 The Department would be grateful for feedback on the options discussed or other suggestions for changes to the methodology of *Port Freight Statistics*, so that we can consider how better to meet user and data supplier needs and how the dissemination of information can be improved.

9.1.2 You can provide feedback by completing a short user feedback survey. Alternatively, you can contact us by **email** to Maritime.Stats@dft.gsi.gov.uk or **phone** 020 7944 4863.

9.2 Specific points for feedback

9.2.1 The Department is particularly interested in your answers to the following questions.

Section 1 – Who you are and why you are interested in these statistics

1. What field do you work in?

   - ☐ Shipping company/ agent
   - ☐ Port operator
   - ☐ Central government or executive agency
   - ☐ Local government
   - ☐ Media – local/ national
   - ☐ Media – transport/ specialist
   - ☐ Transport interest group
   - ☐ University/ researcher
   - ☐ Student
   - ☐ Consultancy
   - ☐ Maritime industry body
   - ☐ Maritime training organisation
   - ☐ Not applicable
   - ☐ Other (please specify)

2. For what purposes do you use *Port Freight Statistics*?

Section 2 – Structure and content of the statistical release

3. How useful do you find the statistical release summary (the PDF document with the text and charts summarising the key figures)?

   ☐ Not at all useful
   ☐ Slightly useful
   ☐ Somewhat useful
   ☐ Very useful
   ☐ Extremely useful
   ☐ I don't read the summary/ Not applicable

   Please elaborate on your answer (e.g. what is useful / not useful?)

4. The technical documentation explains these statistics in more detail, providing information on data collection, classification of goods and vessels, and comparisons with other sources.

   How useful do you find the technical notes?

   ☐ Not at all useful
   ☐ Slightly useful
   ☐ Somewhat useful
   ☐ Very useful
   ☐ Extremely useful
   ☐ I don't read the technical notes/ Not applicable

   Please provide any comments you have on the technical notes.

5. How useful do you find the PORT data tables?

   ☐ Not at all useful
   ☐ Slightly useful
   ☐ Somewhat useful
   ☐ Very useful
   ☐ Extremely useful
   ☐ I don't use the data tables/ Not applicable

   Please provide any comments you have on the data tables (i.e. is there a particular table or set of tables which is of use to you?).

6. Please indicate the extent to which each of the topics covered in the release have proven useful to your work. (Please tick the relevant option for each row) **If you do not use the statistical release summary please move on to Section 3**

	Highly useful	Some use	Little use	No use
Tonnage traffic				
Unitised traffic				
Traffic by direction of travel				
Comparison with other sources				
Cargo category breakdowns				
Port level traffic				
International traffic				
Domestic traffic				
Vessel arrivals				

7. If there are any other topics not covered in the release, but that you would find useful, please list these, and explain how these would be used.

8. Do you have any comments regarding the timing or timeliness of the Port Freight materials that the Department publishes (the statistical release, tables, documentation etc.)?

9. Please provide any further comments you have regarding the structure and content of the statistics series.

If you would like to be added to the *Port Freight Statistics* user email list, please provide your name, organisation and email address at the end of this survey. People on this list will be informed about important updates and be asked their views on topics relating to the statistics series.

Section 3 – Data Supply

10. Do you or your organisation provide *Port Freight Statistics* data to DfT under the Statistical Returns Regulations of 1997 on the Carriage of Goods and Passengers by Sea?

    ☐ Yes
    ☐ No

    If you do not provide data to DfT please proceed to Section 4

11. How long have you been supplying this data to DfT?
    ☐ Less than one year
    ☐ 1 – 2 years
    ☐ 2 – 5 years
    ☐ 5+ years
    ☐ Other (please state)

12. Which method do you use to provide DfT with this data?
    ☐ Internet Statistics Data Entry System (iSDES)
    ☐ Generic Statistical Message (GESMES)
    ☐ ASCII file
    ☐ Paper
    ☐ Other (please state)

13. Do you collect or produce this information for purposes other than to provide us with this information?
    If so, for what other reasons do you collect/produce the data? (e.g. this could be for accounting or billing purposes, or to provide this information to other organisations)

14. Do you carry out any checks on your data before sending it to us? If so, please give details of the checks that you undertake.

15. Do you have any issues when supplying data? If so please provide details of these issues.

16. The Department produces guidance documentation in order to help data suppliers to complete their data returns.

    How useful do you find the guidance documentation?

    - ☐ Not at all useful
    - ☐ Slightly useful
    - ☐ Somewhat useful
    - ☐ Very useful
    - ☐ Extremely useful
    - ☐ I don't read the guidance documentation/ Not applicable

    Please provide any comments you have on the guidance documentation.

17. Is there another method which you would prefer to use to supply data to DfT? (If so please state or explain this method)

18. Would it be possible for you to supply the Bill of Lading or Cargo manifest documents for some or all cargoes for which you are responsible?
    - ☐ Yes, all cargoes
    - ☐ Yes, some cargoes
    - ☐ No
    - ☐ Don't know

    If not, why isn't this possible?

19. Would you like any additional support from DfT regarding supplying data? This could be in the form of training workshops/videos or further guidance materials. If so, please provide details and leave your contact information at the end of this survey.

20. Has DfT or its collecting agent, BMT Reliability Consultants, contacted you to validate your data? If so, how satisfied were you with this experience?
    - ☐ Very satisfied
    - ☐ Somewhat satisfied
    - ☐ Neither satisfied nor dissatisfied
    - ☐ Somewhat dissatisfied
    - ☐ Very dissatisfied

    If you wish, please provide any comments regarding your experiences of these interactions:

Section 4 – Possible changes to methods

Previous chapters of this paper discuss potential changes to the methods used in the current data collection processes.

21. Please indicate the extent to which you think each potential change would be suitable/feasible.

	Very	Somewhat	Not at all	Don't know
Use of sampling – see Section 8				
"Live" validation checks for missing data – Section 6.4.10				
International data collected via another source (e.g. HMRC data) – Section 7.4				

If you wish, please provide further details of why you have indicated your response. Please indicate which response you are referring to.

22. Are there any other data sources which you are aware of which could be used to meet some of our requirements? If so, please state them below.

Section 5 – Contact information

23. Are there any further comments which you would like to raise regarding either the *Port Freight Statistics* publications or data collection methods?

9.2.2 Comments are requested by 22 December 2016.

9.2.3 Please provide your contact details if you do not object to DfT contacting you to discuss your responses further.

Name:
Organisation:
Position:
Telephone:
Email:

9.2.4 Thank you for taking the time to provide a response to our survey. An update on the *Port Freight Statistics* methodology review will be published in March 2017.

9.2.5 Any published conclusions based on responses to this user engagement exercise will be anonymised. However, respondents should be aware that any information provided in response to this review - including personal information - may be subject to publication or release to other parties or to disclosure in accordance with the access to information regimes.[10]

[10] Freedom of Information Act 2000 (FOIA), the Data Protection Act 1998 (DPA) and Environmental Information Regulations 2004.

10 Annex A – Example of validation and grossing methods

10.1.1 Quarterly and annual port traffic data (MDS2 and MSD2X) are received from all major ports. Before these data are validated there are some basic in-built checks which are carried out to check whether the data supplied are in a valid format i.e. the cargo categories are valid, sender ID and other necessary fields are populated

10.1.2 The completeness of the returns from shipping lines/agents (MSD1 forms) are measured by comparing the resulting traffic totals with those from the port (MSD2/2X) returns.

10.1.3 The following example gives an illustration for a fictional port of how the validation procedures, which are carried out for each port, work in practice. Figures are reported for total gross weight (tonnes) of freight traffic into an example port, “Port X”, by cargo category:

Table 2– Reported inwards port traffic (tonnes) data for Port X from ports and shipping agents by cargo category

Cargo Category	MSD1 – quarterly shipping agent port traffic data	MSD2X – Annual port traffic data	Difference
13	693,000	1,064,000	-371,000
19	204,000	239,000	-35,000
21	102,000	194,000	-92,000
29	63,000	102,000	39,000
31	15,000	-	15,000
32	14,000	-	14,000
51	1,000	-	1,000
52	-	-	-
53	-	-	-
54	-	-	-
59	-	-	-
61	124,000	-	124,000
63	19,000	-	19,000
91	1,000	2,000	-1,000
92	37,000	73,000	-36,000
99	282,000	235,000	47,000
Total	1,555,000	1,909,000	-354,000

10.1.4 There are some fairly large discrepancies between cargo category totals here and the overall totals are also quite different. These issues are initially followed up with the port to balance the data and to investigate any missing data. In addition to inwards tonnage (Table 2), similar checks take place for outwards tonnage and units in/out for every major port. At the same time, the cargo category totals are checked against data for the previous year and any differences of +/- 30% are automatically investigated with both the relevant agents and ports.

10.1.5 Further checks are then carried out on the detailed annual port data (MSD2X) against quarterly port data (MSD2) and MSD2X data for the previous year, in order to investigate and understand the discrepancies above.

10.1.6 Quarterly agent data (MSD1) are also checked against the previous year to look for any regular trade routes which may have been missed. The final check which is carried out is called the “mirror check”. This is where the collecting agent looks at domestic (UK port to UK port) data for each port and checks with the corresponding UK port that they have a matching data return. Where they do not match or a return does not exist, this is investigated by contacting the relevant agents and ports.

10.1.7 After detailed conversations with the relevant ports and agents the example contains two sets of figures which are much closer to each other:

Table 3 – Validated inwards port traffic (tonnes) data for Port X from ports and shipping agents by cargo category

Cargo Category	MSD1 – Validated quarterly shipping agent port traffic data	MSD2X – Annual port traffic data
13	991,000	1,064,000
19	198,000	239,000
21	102,000	144,000
29	65,000	102,000
31	15,000	15,000
32	14,000	14,000
51	1,000	2,000
52	-	-
53	-	-
54	-	-
59	-	-
61	124,000	97,000
63	19,000	18,000
91	1,000	2,000
92	35,000	36,000
99	344,000	321,000
Total	1,909,000	2,054,000

10.1.8 Note that the two columns of data resemble each other much more closely, but also that both totals have changed. It is a common outcome that even after this initial validation, the port and agent data still does not match, indicating that there are still errors in either or both sources. It would be extremely resource intensive to continue investigating every difference indefinitely until all data matches, and in most cases would not be possible, hence a further solution is employed.

10.2 Grossing

10.2.1 The annual port traffic data are used to adjust the information supplied by shipping lines and agents, i.e. to estimate for any missing or misclassified data from shipping lines/agents. Annual port traffic data from major ports is deemed more likely to be complete and less likely to contain errors simply because there are far fewer major ports to report data than there are shipping lines and agents.

10.2.2 All of the annual port traffic information provided by each port on the MSD2X form - i.e. the cargo categories for unitised and non-unitised traffic - are divided by the corresponding estimates provided by shipping lines and agents to produce grossing factors. These factors are then applied to all corresponding MSD1 data variables to provide grossed totals. This method allows the estimation of other variables, for example, traffic by cargo type, by port of loading and unloading, vessel characteristics, flag etc., which are not provided by ports. The grossing procedure applies to traffic to and from major ports; information for minor ports is added in separately.

10.2.3 The overall effects of the grossing procedure, and the distribution of grossing factors, are summarised in the following table:

Table 4 – Grossing information for all major UK ports: 2006-2014

	Number of major ports reporting	Total weight reported on MSD1 forms (million tonnes)	Published estimate after grossing (million tonnes)	Implied grossing factor
2006	52	526.5	568.8	1.08
2007	52	521.2	566.6	1.09
2008	52	493.1	548.1	1.11
2009	52	455.8	489.6	1.07
2010	52	463.5	498.5	1.08
2011	51	484.2	507	1.05
2012	51	484.5	489.5	1.01
2013	51	478.9	491.4	1.03
2014	51	474.9	491.5	1.04
2014 results for specific traffic types (major ports)				
Inwards	51	305.9	317.8	1.04
Outwards	51	169.0	173.8	1.03
Liquid Bulk	51	179.1	187.8	1.05
Dry Bulk	51	117.0	121.4	1.04
Other general cargo	51	20.2	21.5	1.06
Lo-lo containers	51	54.9	61.3	1.12
Ro-ro	51	103.7	99.6	0.96
Major ports	51	Smallest implied grossing factor among ports		0.73
	51	Lower quartile (port 13 of 51)		1.00
	51	Median		1.02
	51	Upper quartile (port 39 of 51)		1.07
	51	Largest implied grossing factor among ports		1.21

10.2.4 Returning to the example set out in Tables 2 and 3, grossing factors are calculated by dividing the annual port traffic figures (MSD2X) from major ports by the quarterly shipping line data (MSD1).

Table 5 – Reported inwards port traffic (tonnes) data for Port X from ports and shipping agents by cargo category

Cargo Category	MSD1	MSD2X	Grossing factor
13	991,000	1,064,000	1.07
19	198,000	239,000	1.21
21	102,000	144,000	1.41
29	65,000	102,000	1.57
31	15,000	15,000	1.00
32	14,000	14,000	1.00
51	1,000	2,000	2.00
52	-	-	-
53	-	-	-
54	-	-	-
59	-	-	-
61	124,000	97,000	0.78
63	19,000	18,000	0.95
91	1,000	2,000	2.00
92	35,000	36,000	1.03
99	344,000	321,000	0.93
Total	1,909,000	2,054,000	1.08

10.2.5 Each MSD1 cargo category total will then be grossed to match the corresponding MSD2X total – the relevant grossing factors are listed in table 4 above. This process is followed for each major port to arrive at the final published totals for gross weight in/out and units in/out by cargo category.

11 Annex B – Data collection overview